PAG2ADMG: A Novel Methodology to Enumerate Causal Graph Structures

Nishant Subramani and Doug Downey
Northwestern University, Evanston IL 60201, USA
nishantsubramani2017@u.northwestern.edu
ddowney@eecs.northwestern.edu

Abstract

Causal graphs, such as directed acyclic graphs (DAGs) and partial ancestral graphs (PAGs), represent causal relationships among variables in a model. Methods exist for learning DAGs and PAGs from data and for converting DAGs to PAGs. However, these methods only output a single causal graph consistent with the independencies/dependencies (the Markov equivalence class M) estimated from the data. However, many distinct graphs may be consistent with M, and a data modeler may wish to select among these using domain knowledge. In this paper, we present a method that makes this possible. We introduce PAG2ADMG, the first method for enumerating *all* causal graphs consistent with M, under certain assumptions. PAG2ADMG converts a given PAG into a set of acyclic directed mixed graphs (ADMGs). We prove the correctness of the approach and demonstrate its efficiency relative to brute-force enumeration.

Introduction

Capturing causal relationships is essential for a comprehensive model of a domain. However, it is difficult to identify causal relationships from observational data because many distinct causal models may correspond to the same associative relationships discovered from data. In such a situation, a data modeler or domain expert may wish to enumerate *all* the causal graphs consistent with the probabilistic independencies and dependencies present among variables of a dataset, in order to (for example) leverage domain knowledge to choose one graph from the consistent set.

Consider a physician in a major hospital who has observed symptom X among many of her patients. Based on their logs, patients developing symptom X had a much higher likelihood of mortality. To understand the underlying causes of symptom X, the physician aims to learn a causal model from her data. We consider causal models that are graphs, where nodes represent random variables and edges denote probabilistic relationships between variables. Methods exist to learn models such as directed acyclic graphs (DAGs) and partial ancestral graphs (PAGs) from data (Spirtes, Glymour, and Scheines 2000). Using these methods, the physician can learn one model that describes a set of independence relationships between variables of interest (a Markov equivalence class), including symptom X.

Although DAGs and PAGs describe causal relationships, these models fail to capture the full set of causal models in a specific Markov equivalence class (Richardson 2003). Acyclic Directed Mixed Graphs (ADMGs), on the other hand, provide richer representations: the full set of ADMGs in a specific Markov equivalence class provides an explicit enumeration of the entire causal space.

In this paper, we present a method that enumerates, for a given PAG (or DAG converted into a PAG), the full set of ADMGs in the PAG's Markov equivalence class. Equipped with the complete set of ADMGs, the physician could use domain knowledge to eliminate structures that have causal fallacies, use model-averaging to build better classifiers, or simply choose a single preferred model from the set. Thus, our method provides a data modeler novel and potentially valuable model selection choices.

Partial Ancestral Graphs

A PAG $\mathscr{P}$ is a causal graph that uniquely represents one Markov equivalence class M of maximal ancestral graphs (MAGs), a strict subset of ADMGs. MAGs are a special case of mixed graphs that must satisfy the ancestral and maximal properties. Only one of four different edges can exist between a pair of nodes in $\mathscr{P}$: tail-arrow (directed, $\rightarrow$), arrow-arrow (bi-directed, $\leftrightarrow$), circle-arrow (o$\rightarrow$), and circle-circle (o–o) (Zhang 2008). A tail or arrow mark appears at a node X in $\mathscr{P}$ if and only if every MAG in M contains this mark at X. A circle mark denotes a node at which two MAGs in M have differing marks. M-separation is a property of causal mixed graphs that helps to determine independence relationships between nodes. Due to space constraints, we refer to (Zhang 2008) for more details on MAGs and m-separation.

Arrows on edges represent the direction of causal influence. The presence of an edge in $\mathscr{P}$ between nodes X and Y indicates that there exists no set of nodes **Z**, including the empty set, that m-separates X and Y (Richardson and Spirtes 2002). $X \rightarrow Y$, indicates that X is a causal ancestor of Y and Y is not an ancestor of X. $X \leftrightarrow Y$, indicates that X is not an ancestor of Y and Y is not an ancestor of X. X o$\rightarrow Y$, indicates that X is not an ancestor of Y.

Acyclic Directed Mixed Graphs

ADMGs are a specific type of mixed graph that contain only directed and bi-directed edges and have no directed cycles in the absence of selection variables. ADMGs do not have to satisfy ancestral or maximal conditions and thus are a much less restrictive super-set of MAGs. They are also trivially a strict super-set of DAGs. Furthermore, between any pair of nodes, both a directed and bi-directed edge can occur, granting additional expressive power (Richardson 2003).

Theorems

Theorem 1 *Every ADMG $\mathscr{A}$ in the Markov equivalence class M has the same skeleton of PAG $\mathscr{P}$ describing M.*

Theorem 2 *Let $\mathscr{A}$ be an ADMG for M and let PAG $\mathscr{P}$ describe M. Then for every pair of nodes X and Y:*

1. *if $X \rightarrow Y \in \mathscr{P}$, then $X \rightarrow Y \in \mathscr{A}$ and $X \leftrightarrow Y \notin \mathscr{A}$*
2. *if $X \leftrightarrow Y \in \mathscr{P}$, then $X \rightarrow Y \notin \mathscr{A}$ and $Y \rightarrow X \notin \mathscr{A}$*
3. *if X o→ $Y \in \mathscr{P}$, then $Y \rightarrow X \notin \mathscr{A}$*
4. *$X \rightarrow Y$ and $X \leftrightarrow Y \in \mathscr{A}$ if and only if there exist two ADMGs, $\mathscr{A}_1$ and $\mathscr{A}_2$ for M, with exactly the same structure as $\mathscr{A}$ apart from this edge such that $X \rightarrow Y \in \mathscr{A}_1$ and $X \leftrightarrow Y \in \mathscr{A}_2$.*

Corollary 2.1 *$X \rightarrow Y$ and $X \leftrightarrow Y \in \mathscr{A}$ if and only if X o–o $Y \in \mathscr{P}$ or X o→ $Y \in \mathscr{P}$. (Follows from Theorem 2.4)*

We believe the results to be novel in this work.

PAG2ADMG

The PAG2ADMG method transforms a PAG $\mathscr{P}$, which describes a Markov equivalence class M, into the full set of ADMGs **S** for M. This method assumes that there are no selection variables and contains a set of generation steps (1-3) and two pruning steps (4-5).

Input: A PAG $\mathscr{P}$ over a set of variables **V** that describes a Markov equivalence class M
Output: The set of all ADMGs **G** over the same set of variables **V** in M

1. Every edge $E \in \mathscr{P}$ between each pair of nodes $X, Y \in$ **V** is converted into the set of all possible edges **E*** between X, Y any ADMG $\mathscr{A} \in M$ could have. (Uses Theorems 1-2 & Corollary 2.1)
2. All possible graphs **G** are generated using all combinations of edge possibilities **E*** between each pair $X, Y \in$ **V**.
3. All graphs with directed cycles are removed (only acyclic graphs remain).
4. Each ADMG $\mathscr{A}^*$ is converted to a DAG $\mathscr{D}^*$ by replacing each bi-directed edge with a latent variable confounder.
5. Each DAG $\mathscr{D}^*$ is then converted to a PAG $\mathscr{P}^*$ using a method from (Richardson and Spirtes 2002). If $\mathscr{P}^* \not\equiv \mathscr{P}$, then **G** $\setminus \mathscr{A}^*$.

Example

The PAGs in Fig. 1 and Fig. 3 represent different Markov equivalence classes. The circle-arrow edges in Fig. 1 and Fig. 3 result in the possibility of of a directed edge, bi-directed edge, or both edges being present in a Markov equivalent ADMG. For these two examples, each combination of those three types of edges for both occurrences of the circle-arrow edge defines a unique Markov equivalent ADMG. Thus each PAG is described by 9 different distinct ADMGs, examples of which are shown in Fig. 2 and Fig. 4 respectively. PAG2ADMG reduces the number of generated ADMGs from $6^{\binom{4}{2}} = 46656$ using a brute-force method to 32 and 16 when converting Fig. 1 and Fig. 3 respectively.

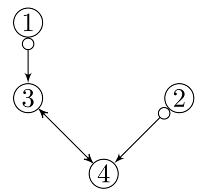

Figure 1: PAG Over 4 nodes (Richardson 2003)

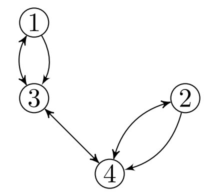

Figure 2: 1 of 9 Markov Equivalent ADMGs

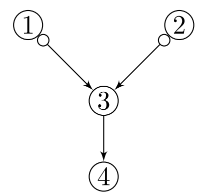

Figure 3: Y-PAG Over 4 Nodes

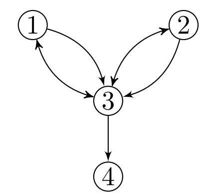

Figure 4: 1 of 9 Markov Equivalent ADMGs

Conclusion

This work[1] makes the following contributions:

1. Provides an algorithm to transform a PAG to its Markov equivalent set of ADMGs over those same variables.
2. Provides proof of correctness for the method.[2]
3. Provides code & documentation to run PAG2ADMG. [2]

References

Richardson, T., and Spirtes, P. 2002. Ancestral graph Markov models. *Annals of Statistics* 962–1030.

Richardson, T. 2003. Markov properties for acyclic directed mixed graphs. *Scandinavian Journal of Statistics* 30(1):145–157.

Spirtes, P.; Glymour, C. N.; and Scheines, R. 2000. *Causation, prediction, and search*. MIT press.

Zhang, J. 2008. On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence* 172(16):1873–1896.

[1] The work was supported in part by NSF Grants IIS-1640380 and IIS-1351029.

[2] Can be found at `github.com/hatat5/PAG2ADMG`